DOCUMENT RESUME

ED 114 750 CG 010 211

AUTHOR Jackson, Douglas N.
TITLE Discriminantly Valid Personality Measures: Some Propositions. Research Bulletin No. 339.
INSTITUTION University of Western Ontario, London. Dept. of Psychology.
REPORT NO ISSN-0316-4675; R Bull-339
PUB DATE Sep 75
NOTE 15p.

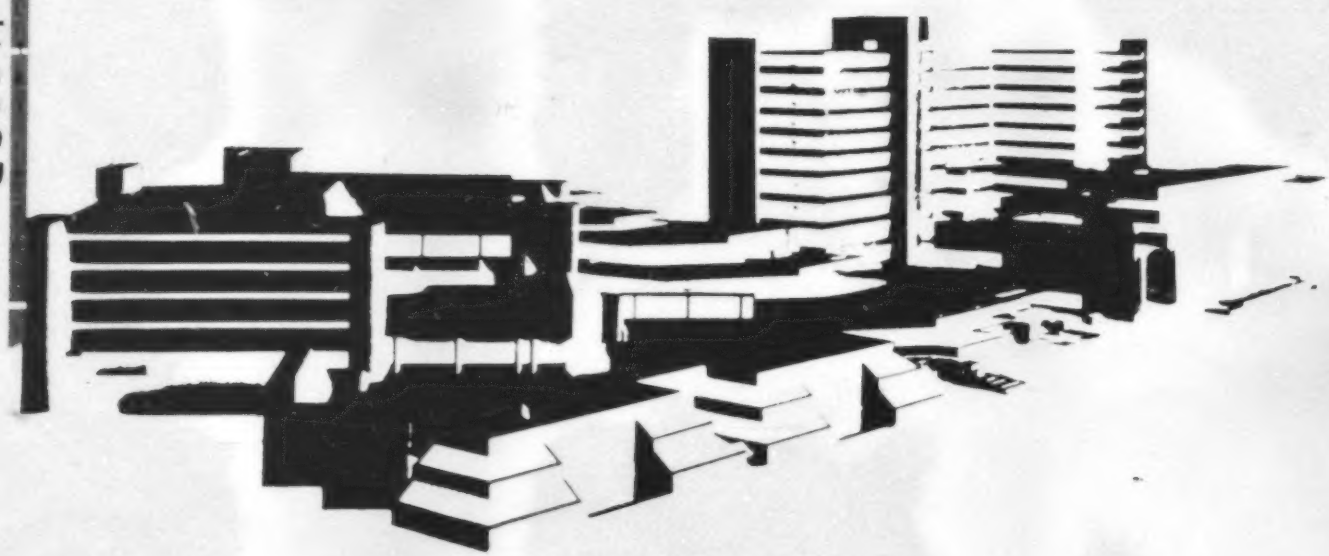
EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Discriminant Analysis; *Measurement; *Personality Assessment; Personality Theories; *Predictive Validity; Psychometrics; Situational Tests; Symposia; *Test Validity

ABSTRACT

Starting with the premise that the construct-oriented approach is the only viable approach to personality assessment, this paper considers five propositions. First, a prerequisite to generalizable and valid psychometric measurement of personality rests on the choice of broad-based constructs with systematic univocal definitions. Next, measures will be likely to demonstrate convergent and discriminant validity and relative freedom from desirable bias if, and only if, these requirements are incorporated into test construction. Third, untrained item writers produce more valid personality scales than do empirical procedures using extant pools. Fourth, convergent and discriminant validity of personality measures is demonstrable even in the face of confounded method variance, and last, a potentially fruitful avenue for further investigation is the demonstration of a number of reproducible sets of psychopathological types. Evidence supporting each of the above propositions is discussed. Some areas of research needing attention are outlined. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED114750



DISCRIMINANTLY VALID PERSONALITY MEASURES: SOME PROPOSITIONS[1]

Douglas N. Jackson

University of Western Ontario

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Research Bulletin No. 339
September, 1975

ISSN-0316-4675

[1]Presented at a symposium on The Future of Personality Assessment at the meeting of the American Psychological Association, Chicago, August 30, 1975. Supported in part by a research grant from Canada Council.

DEPARTMENT OF PSYCHOLOGY

THE UNIVERSITY OF WESTERN ONTARIO LONDON, CANADA

Discriminantly Valid Personality Measures: Some Propositions[1]

Douglas N. Jackson

University of Western Ontario

In 1957--18 years ago--at an APA convention I gave a paper--co-authored by Messick--entitled Content and Style in Personality Assessment. This paper was interpreted as advancing the hypothesis that the major response dimensions of some widely used personality tests were identifiable as response styles. But the paper carried another message, one that is even more important for personality assessment than that of response styles. It is the idea that personality assessment involves the appraisal of content, referring "to response consistencies in certain defined assessment situations which reflect a particular set of broader behavioral tendencies, relatively enduring over time, having as [its] basis some unitary trait, need state, attitudinal or belief disposition, or psychopathological syndrome" (Jackson & Messick, 1962). The suggestion that what is to be measured in personality assessment should reflect content, so defined, ought not to be very revolutionary, were it not for a tradition of radical empiricism which left most personality assessment specialists unable and/or unwilling to admit to an interest in measuring content and markedly suspicious of anyone else's claim of having done so. The state of knowledge in personality and psychopathology, we were told, was not so advanced as to permit unequivocal identification of the relationship between responses to item content and the underlying dynamics of trait dimensions.

Furthermore, when investigators such as Campbell and Fiske (1959) sought to evaluate the degree to which putative trait measures were generalizable across methods of measurement, the results were almost uniformly discouraging. It is little wonder that many psychologists dispaired of the possibilities for valid and socially useful measurement in personality.

But the emergence of the thinking behind the notion of construct validity and the idea, championed by Loevinger (1957), that there should be some theoretically defensible relationship between item content and the underlying construct, provided an important alternative to the radical empiricism implicit in building personality scales out of items found to discriminate particular criterion groups. What would happen if personality scales were developed on the basis of a frankly rational approach which created an item pool by writing items so that they spanned all identifiable facets of an explicit definition of a trait, including various likely manifestations and evoking situations. In 1960 I began a program of personality scale research and construction in which I have had the opportunity to review more than 15 thousand items. Much in the way of cumulative knowledge about personality assessment has emerged from this experience. One of the strongest impressions to emerge from this experience is in the potency of human judgment of item content, and in the importance of basing personality assessment upon well-thought-out constructs. Thus, my first proposition is that

> A prerequisite to generalizable and valid psychometric measurement of personality dimensions rests on the choice of broad-based constructs with systematic univocal definitions.

Notice that I did not equivocate and say 'one approach to valid measurement is....' Rather, I have implied that this is the only viable approach. The alternative of, for example, the use of narrow, situation specific constructs is at best inefficient, in that it will require a very large number of scales for situation-trait facets, and at worst will fail entirely because very little in the way of systematic theory is available about how situations should be categorized and how precisely they may interact with traits.

There are times when one may with impunity move from one level of abstraction to another. For example, Jackson, Hourany and Vidmar (1972) undertook a conceptual analysis of the construct of Risk Taking. They hypothesized four facets: Monetary, Physical, Social, and Ethical Risk Taking. They undertook to measure each facet with five different methods of measurement. Factor analytic treatment of the data revealed four distinct factors, representing the four facets of risk taking. The pattern of factor loadings was virtually completely consistent with expectations regarding the convergent and discriminant validity of the four facets. However, when a second-order factor analysis was undertaken, clear evidence for a general Risk Taking factor was uncovered, accounting for 80 per cent of the first-order factor variance. In the case of Risk Taking, one could do reasonably well predicting behavior relevant to sub-species of Risk Taking by using a general measure.

The situation is different for Achievement, long regarded as a unitary dimension. Jackson, Ahmed, and Heapy (1976) undertook a conceptual analysis of Achievement and hypothesized six distinct facets: Status with Experts, Acquisitiveness, Achievement via Independence, Status with Peers,

Competitiveness, and Concern for Excellence. Five methods of measuring each of these facets were devised and administered to a large group of subjects. Again factor analytic results revealed a pattern of factor loadings highly consistent with convergent and discriminant properties for each of these six facets. This indicates that the achievement construct as described by McClelland and others is actually separable into distinct components each of which can be isolated in psychological measures. Even when a further factor analysis was undertaken of the first-order factor intercorrelation matrix, three factors emerged, indicating that even at a higher level of abstraction there was no confirmation of a single dimension of Achievement.

Results such as those reported indicating evidence for convergent and discriminant validity are typical of many similar findings obtained using personality inventories that we have developed. In contrast to the experience of many other investigators, our relative success is, I submit, the result of our approach to test construction. Our suggestions for constructing personality assessment devices are fairly simple, as suggested by the second proposition:

> Measures will be likely to demonstrate convergent and discriminate validity and relative freedom from desirability bias if, and only if, these requirements are incorporated into test construction.

To construct a personality assessment device, one need only select an appropriate dimension, think about it in terms of its manifestations and in terms of its relation to other constructs, carefully define it, and prepare an item pool. Items should be balanced in terms of likely

manifestations of the trait, and in terms of negative and positive instances. Even if one is preparing only a single scale, it is better to consider items in relation to a multi-scale battery. This is so because it forces one to consider not only the item's substantive link to its own scale, but its differentiation from distinct scales. Items should have a strong conceptual link to only one scale.

We have been taught to be suspicious of such heavy reliance on rationality. Is it likely that a psychologist can simply look at an item and determine what scale it is keyed on? Our experience is illustrative. An analysis was undertaken (Jackson, 1971) with the item pool comprising the Personality Research Form, and in particular the degree to which an item written for a particular scale correlated highly with an irrelevant scale. In the entire set of over 900 items, only five failed in this respect, although, to be sure, more were found wanting in other respects. The percentage of hits using this criterion was 99.9. This is strong evidence, I think, that substantive analysis alone can contribute much to the discriminant properties of scales.

The recommended procedure is to prepare a matrix of correlations between each item and each of a substantial number of irrelevant scales, including a desirability scale. When an item correlates too highly with an irrelevant scale, discard the item. Actually this procedure can be refined by partialling desirability variance out of the variance associated with both item and total scale score, and through the use of orthogonal factor scores rather than raw scale scores. We have also developed algorithms to reduce scale intercorrelations.

In general, what I am suggesting is that if items are selected to be associated with only their own scale, if scales are constituted so they will have only minimum mutual redundancy, if items bear an important substantive relation to a scale definition, if response biases are suppressed in scale construction, a likely result is the development of scales showing worthwhile levels of convergent and discriminant validity. I have with my collaborators completed about ten studies of convergent and discriminant validity involving a number of different sets of scales--all but one yielded confirmatory evidence. The single exception was when we attempted to have prison guards judge the personalities of their wards. They seemed to be validly sensitive to only one dimension--that of Hostility.

The approach I am advocating places a heavy reliance on judgment. We have relied on the judgment of those relatively experienced in personality and psychological theory. The question arises as to whether or not this is a necessary precondition for obtaining the sort of results that I have reported. I think not. In fact, judging from the overall history of personality scale construction, psychological training might even be a disadvantage. I believe that persons of hardly more than average sophistication in psychology can draw valid inferences about personality items. Let me give you an example of a multidimensional scaling analysis based upon two independent groups of Canadian Forces officer candidates. We asked these officer candidates to judge the mutual relations between persons, between items, and between persons and items, for example, by asking them to judge the probability that a certain person, whose description was provided, would respond true to a given item. When our subjects were split randomly into two groups, and entirely independent, multidimensional scaling analyses were conducted on each sample, it was

found that all items and persons showed highest projections on the appropriate hypothesized dimension of Dominance, Autonomy, or Impulsivity. These scale values for items and persons replicated across samples substantially, as indicated by correlations of .99, .99, and .98, respectively.

Now, let's turn to Proposition 3.

Untrained item writers produce more valid personality scales than do empirical procedures using extant pools.

In 1971 (Jackson, 1971) I issued a challenge to investigators to undertake such a comparison. Ashton and Goldberg (1973) accepted the challenge and uncovered considerable support for the proposition above. My own study (Jackson, 1975) evaluated it further. What I did was to select three personality scales--Tolerance, Sociability, and Self Esteem--and assign one of them randomly to each of 23 students in an undergraduate course with instructions to write 16 items relevant to a definition given. We administered these student-constructed scales together with similar scales drawn from the Jackson Personality Inventory and the California Psychological Inventory. The latter device is considered by some psychologists to be the finest example of empirical scale construction. Validity was evaluated by computing correlations between each of the experimental and comparison scales and self- and peer-rating criteria based on 116 females comprising roommate pairs. Data revealed that the magnitude of the average validities of student-constructed scales was about three times higher than those derived from the CPI.

I would like to suggest that the basis on which relatively novice persons can accurately judge and write items is by using a shared trait inferential network. This is conceptualized as an n-dimensional Euclidean space in which traits are organized in terms of their implicative relations, which in turn are a function of their actual co-occurrence. Recent evidence

for such a shared network was developed by Reed and Jackson (1975), who demonstrated that groups of judges showed correlations in excess of .98 in ascribing personality scale responses to certain psychopathological types, descriptions of which they had been given. Furthermore, there was a strong evidence that their judgments were accurate in reflecting actual responses.

Once relatively homogeneous scales of personality of psychopathology have been identified, the question arises as to whether or not one can identify cluster of persons showing unique patterns of high and low scores. Our experience from a number of analyses is that such unique profiles are identifiable and can be replicated. Let me give an illustration. Skinner, Jackson, and Hoffmann (1974) administered the Differential Personality Inventory to psychiatrically-hospitalized alcoholics and decomposed the data matrix in such a way that similarities in profile shape were identified. The criterion for identifying a conman modal profile type was that it be replicated across three sub-samples. Eight such bipolar modal profile types were found, each quite different from the average profile for all alcoholics. Further studies have revealed that the proportion of deviant types varies in different normal and pathological samples. A fruitful avenue for investigation would be the use of types so isolated in studies of prognosis and the differential effects of treatment.

The Future of Personality Assessment

What does the future hold for personality assessment? Rather than prognosticate regarding the possible course of developments in personality assessment, let me simply list a few of the areas in which I believe research is urgently needed.

First, in regard to the study of situations, how they influence behavior and interact with traits, and for that matter, how persons influence situations, we are in a vast sea of ignorance. Although we have

been reminded frequently of the importance of situational determinants at least since the 1935 Murchison Handbook of Social Psychology, and although hard data are frequently promised, they are rarely delivered. Rarely have there been attempts to classify and measure situational contexts parallelling the many attempts at identifying the important dimensions of personality. It is, of course, very difficult and expensive to devise life-like situational contexts for assessment purposes. One promising approach is the use of verbally elicited responses to materials having the demonstrated tendency to elicit a certain class of behavior. It is possible to treat large numbers of such items in the same analytical and mathematical ways in which we treat personality items, providing a basis for the emergence of valid generalizations.

Secondly, the methods by which we validate personality assessment devices need further investigation. In a sense, a multitrait-multimethod matrix can be looked upon as an evaluation of the degree to which personality traits are generalizable over situations. But how much is known about the typical set of criterion measures? The usual procedure is to spend several years carefully devising a personality questionnaire, but only a few minutes with our criteria. Of course, one way to obtain a handle on the identification of uncontaminated trait variance is to seek to understand the sources of method variance which inevitably contaminate our criteria.

Third, we should seek new techniques for measuring personality, apart from the printed questionnaire. I am referring to techniques which are structured and possibly even contrived, but which do not require a self report. The Internation-Simulation of Guetzkow and Cherryholmes (1966), in which individuals play the role of government leaders in prescribed situations, strikes me as promising for allowing the emergence of personal qualities.

Possibly other kinds of similar simulations in which a computer terminal is an integral part may emerge.

About one thing I am relatively certain. Personality assessment has outgrown its dependence on ad hoc and theoretically vacuous techniques such as empirical scale construction with respect to external criteria. It is important to know something psychologically about what we are measuring. I for one have confidence that we shall continue to make progress measuring psychologically important variables.

References

Ashton, S. G., & Goldberg, L. R. In response to Jackson's challenge: The comparative validity of personality scales constructed by the external (empirical) strategy and scales developed intuitively by the experts, novices, and laymen. Journal of Research in Personality, 1973, 7, 1-20.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation of the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Guetzkow, H., & Cherryholmes, C. H. Internation Simulation Kit. Chicago: Science Research Associates, 1966.

Jackson, D. N. The dynamics of structured personality tests: 1971. Psychological Review, 1971, 78, 229-248.

Jackson, D. N. The validity of personality scales constructed by naive item writers. Educational and Psychological Measurement, 1975, 35, 361-370.

Jackson, D. N., Ahmed, S., & Heapy, N. Is achievement a unitary construct? Journal of Research in Personality, 1976, in press.

Jackson, D. N., Hourany, L., & Vidmar, N. J. A four-dimensional interpretation of risk taking. Journal of Personality, 1972, 40, 483-501.

Jackson, D. N., & Messick, S. Content and style in personality assessment. Psychological Bulletin, 1958, 55, 243-252. Reprinted in the Bobbs-Merrill Series in the Social Sciences. In 29-9 Sarason (Ed.), Contemporary research in personality. Princeton, N. J.: Van Nostrand, 1962, 16-23.

Jackson, D. N., & Messick, S. Response styles and the assessment of psychopathology. In S. Messick and J. Ross (Eds.), Measurement in personality and cognition. New York: Wiley, 1962. Reprinted in D. N. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967, Pp. 541-558.

Loevinger, J. Objective tests as instruments of psychological theory. Psychological Reports, 1957, 3, 635-694.

McClelland, D. C. The achieving society. Princeton, New Jersey: Van Nostrand, 1961.

Reed, P., & Jackson, D. N. Clinical judgment of psychopathology: A model for inferential accuracy. Journal of Abnormal Psychology, 1975, in press.

Skinner, H. A., Jackson, D. N., & Hoffmann, H. Alcoholic personality types: Identification and correlates. Journal of Abnormal Psychology, 1974, 83, 658-666.

Footnote

[1]Presented at a symposium on the Future of Personality Assessment at the meeting of the American Psychological Association, Chicago, August 30, 1975. Supported in part by a research grant from Canada Council.